NOAA Earth System
Research Laboratory

A comparison of calibrated T_{2m} probabilistic forecasts from GFS and ECMWF reforecasts

Tom Hamill and Jeff Whitaker

NOAA / ESRL / PSD, Boulder, CO

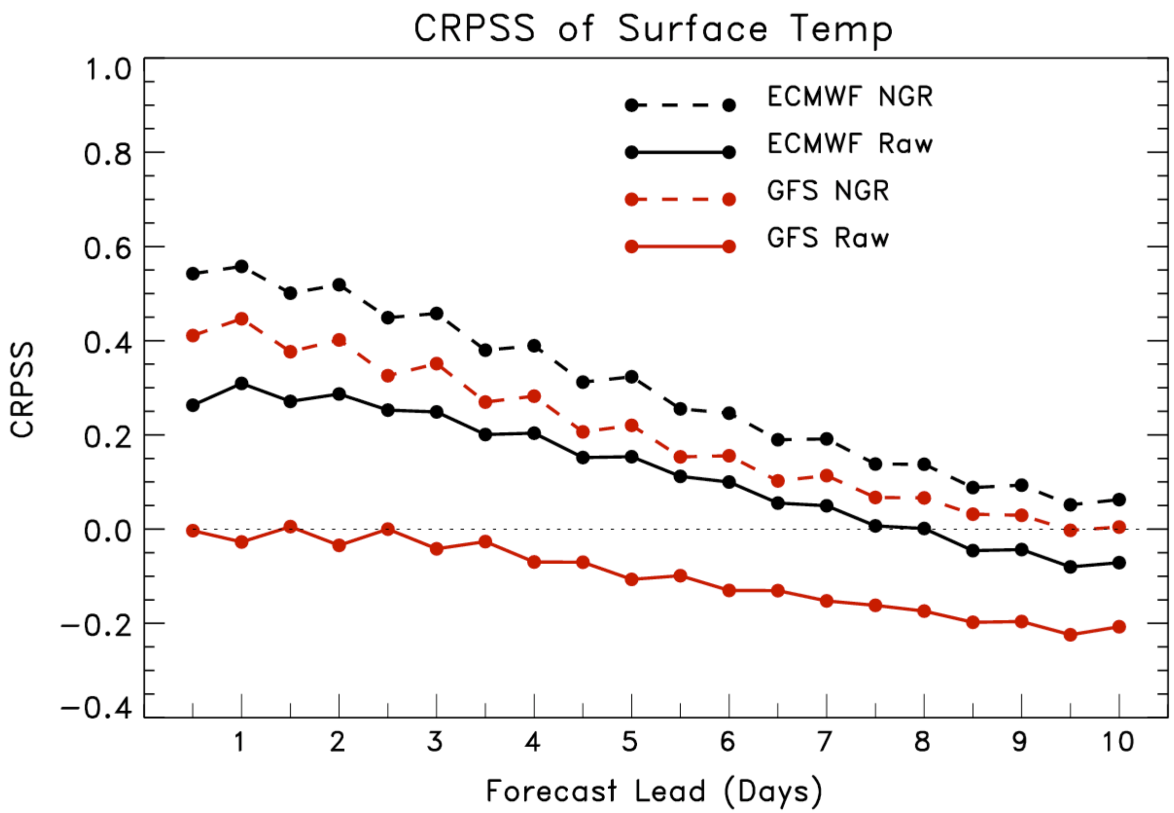
Renate Hagedorn

ECMWF, Reading, England

Bottom-line messages

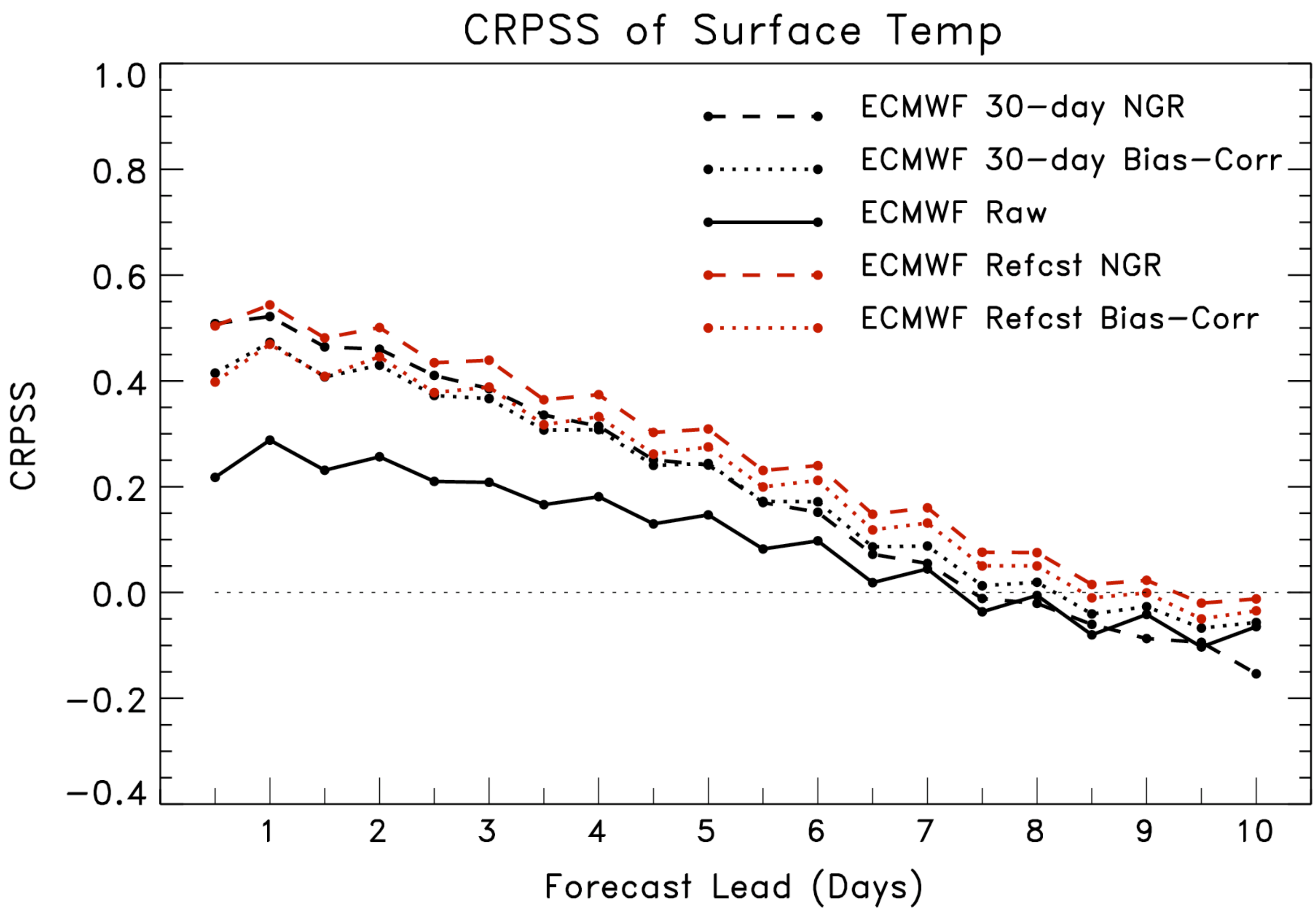
(1) Calibrated GFS based on 1998 ensemble more skillful than probabilities from raw ECMWF ensemble.

(2) Substantial improvement of ECMWF ensemble based on reforecasts; smaller amount than GFS, but still large.



Bottom-line messages

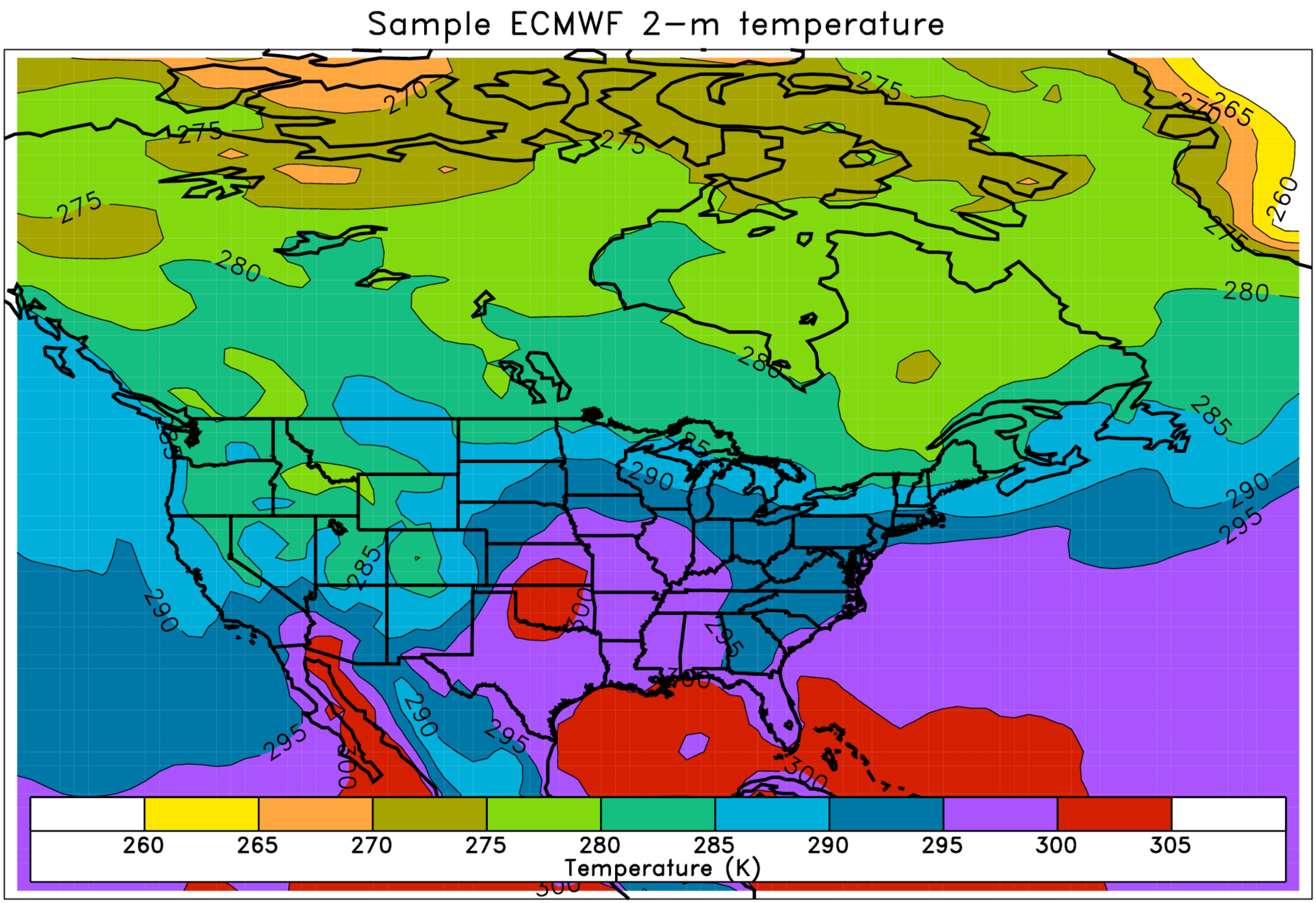
(3) 30-day bias corrections do a good job of correcting short-term forecasts. Somewhat less useful in medium range.



ECMWF's reforecast data set

- **Model**: 2005 version of ECMWF model; T255 resolution.
- **Initial Conditions**: 15 members, ERA-40 analysis + singular vectors
- **Dates of reforecasts**: 1982-2001, Once-weekly reforecasts from 01 Sep - 01 Dec, 14 total. So, 20*14 ensemble reforecasts = 280 samples.
- **Data** sent to NOAA / ESRL : T_{2M} ensemble over most of North America, excluding Alaska. Saved on 1-degree lat / lon grid. Forecasts to 10 days lead.

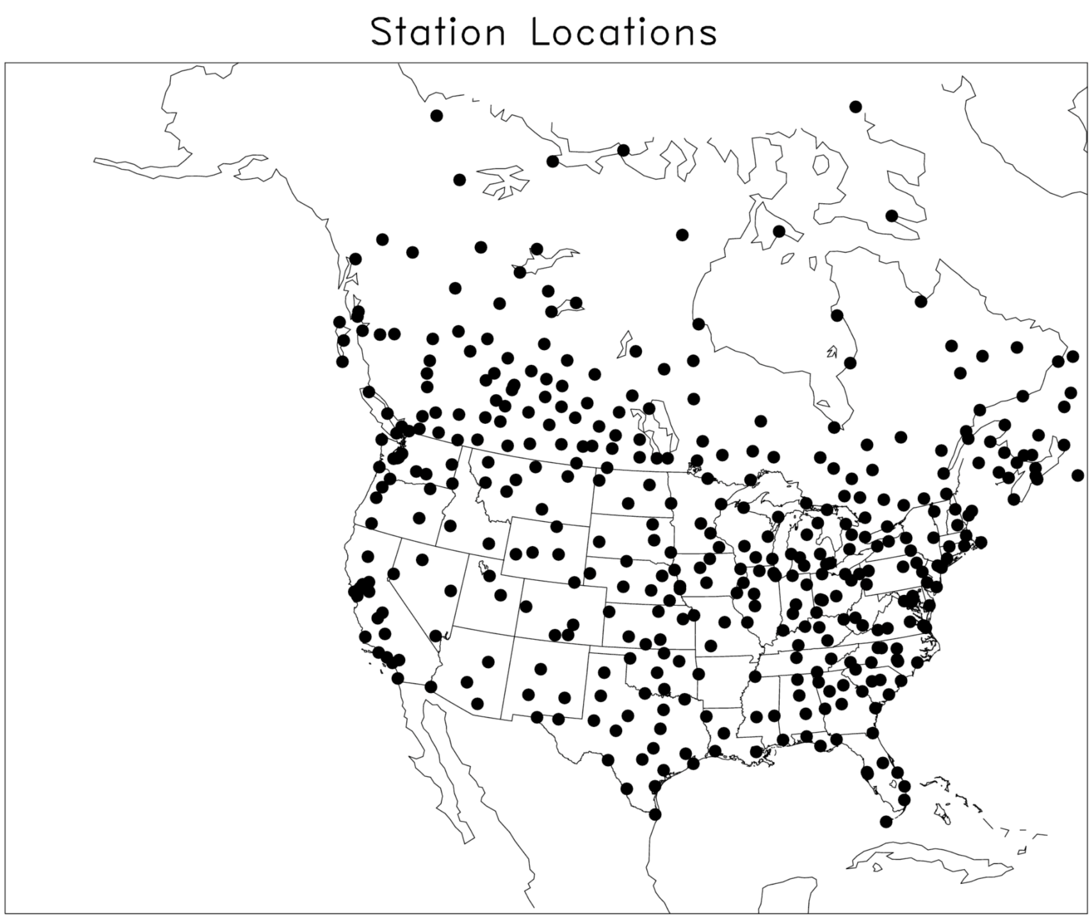
ECMWF domain sent to us for reforecast tests



NOAA's reforecast data set

- **Model**: T62L28 NCEP GFS, circa 1998
- **Initial States**: NCEP-NCAR Reanalysis II plus 7 +/- bred modes.
- **Duration**: 15 days runs every day at 00Z from 19781101 to now. (*http://www.cdc.noaa.gov/people/jeffrey.s.whitaker/refcst/week2*).
- **Data**: Selected fields (winds, hgt, temp on 5 press levels, precip, t2m, u10m, v10m, pwat, prmsl, rh700, heating). NCEP/NCAR reanalysis verifying fields included (Web form to download at *http://www.cdc.noaa.gov/reforecast*). Data saved on 2.5-degree grid.
- Here, use only the subset of data overlapping with ECMWF reforecast data set.

Observation Locations



Uses stations from NCAR's DS472.0 database that have more than 96% of the yearly records available, and overlap with the domain that ECMWF sent us.

Calibration Procedure: "NGR"
"Non-homogeneous Gaussian Regression"

- **Reference**: Gneiting et al., *MWR*, **133**, p. 1098
- **Predictors**: ensemble mean and ensemble spread
- **Output**: mean, spread of calibrated normal distribution

$$f^{CAL}\left(\bar{\mathbf{x}}, \sigma\right) \sim N\left(a + b\bar{\mathbf{x}}, c + d\sigma\right)$$

- **Advantage**: leverages possible spread/skill relationship appropriately. Large spread/skill relationship, c ≈ 0.0, *d* ≈1.0. Small, *d* ≈ 0.0
- **Disadvantage**: iterative method, slow…no reason to bother (relative to using simple linear regression) if there's little or no spread/skill relationship.

Training Data for
Non-homogeneous Gaussian Regression
(all cross validated)

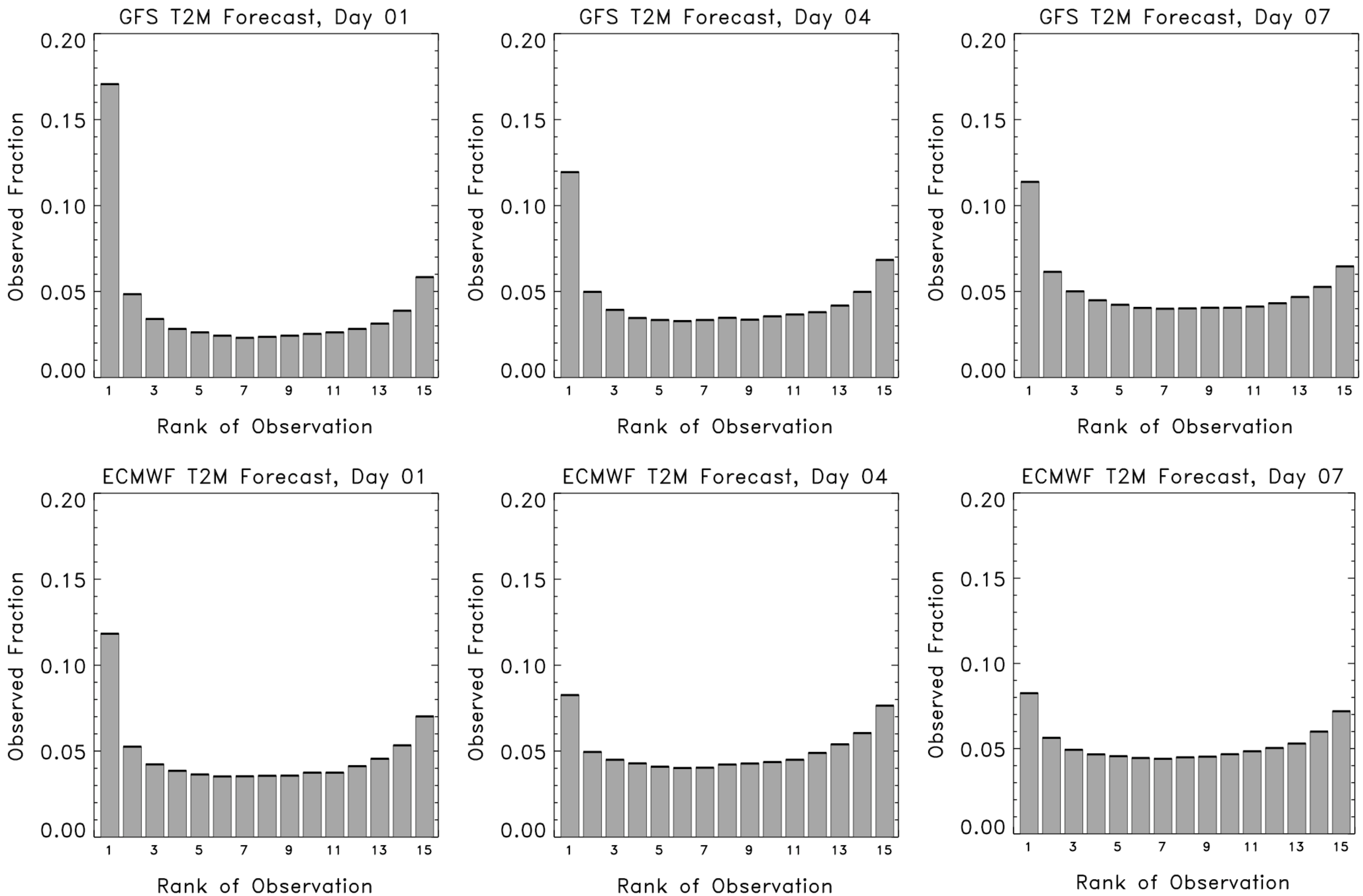
- **01 Sep**: *01 Sep*, 08 Sep, 15 Sep
- **08 Sep**: 01 Sep, *08 Sep*, 15 Sep, 22 Sep
- **15 Sep**: 01 Sep, 08 Sep, *15 Sep*, 22 Sep, 29 Sep
-
-
-
- **17 Nov**: 03 Nov, 10 Nov, *17 Nov*, 24 Nov, 01 Dec
- **24 Nov**: 10 Nov, 17 Nov, *24 Nov*, 01 Dec
- **01 Dec**: 17 Nov, 24 Nov, *01 Dec*

Use a centered training data set for weeks 3 - 12, uncentered for weeks 1, 2, 13, and 14

… but first, rank histograms

GFS

ECMWF



Members randomly perturbed by 1.0K to account for observation error; probably a bit small for GFS on its coarser 2.5º grid, which would make their histograms slightly more uniform. Ref: Hamill, *MWR*, **129**, p. 556.

Continuous Ranked Probability Score (CRPS) and Skill Score (CRPSS)

$$CRPS^f_{i,j,k} = \int_{-\infty}^{+\infty} \left[F_{i,j,k}(y) - F^o_{i,j,k}(y) \right]^2 dy$$

$i = 1, \ldots, \#\ case\ days$

$j = 1, \ldots, \#\ years\ of\ reforecasts$

$k = 1, \ldots, \#\ station\ locations$

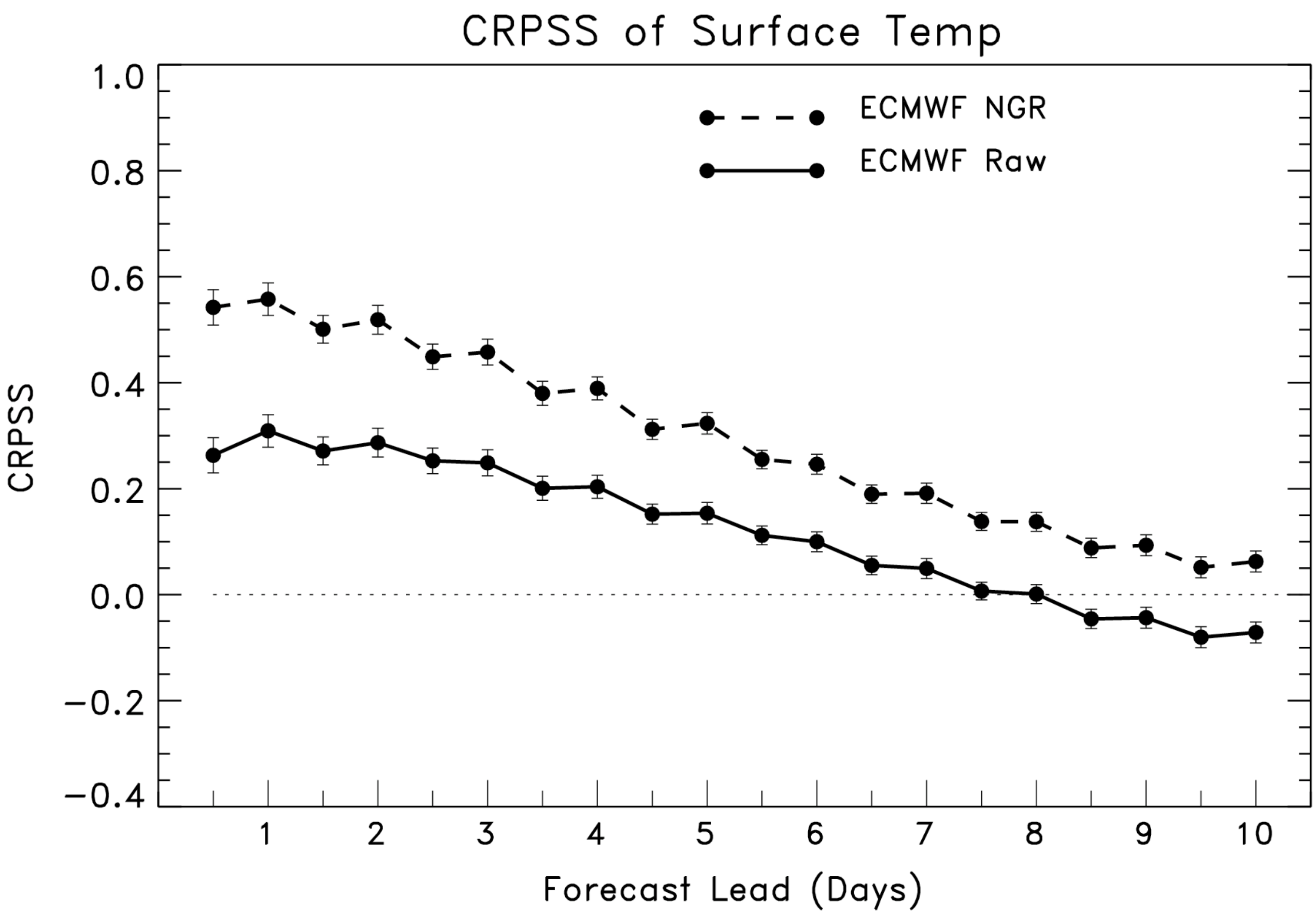
$F_{i,j,k}(y)\ is\ forecast\ CDF\ at\ value\ y$

$F^o_{i,j,k}(y)\ is\ obs\ CDF\ at\ value\ y\ (Heaviside)$

$$CRPSS = 1.0 - \frac{\overline{CRPS}^f}{\overline{CRPS}^c}$$

(This conventional way of calculating CRPSS exaggerates skill if some samples have more climatological spread than others. Will use a modified version where we calculate CRPSS separately for 8 different categories of climatological spread and then average them. See Hamill and Juras, January 2007, *QJRMS,* and Hamill and Whitaker (2007) *MWR,* to appear, tinyurl.com/29oy8s)

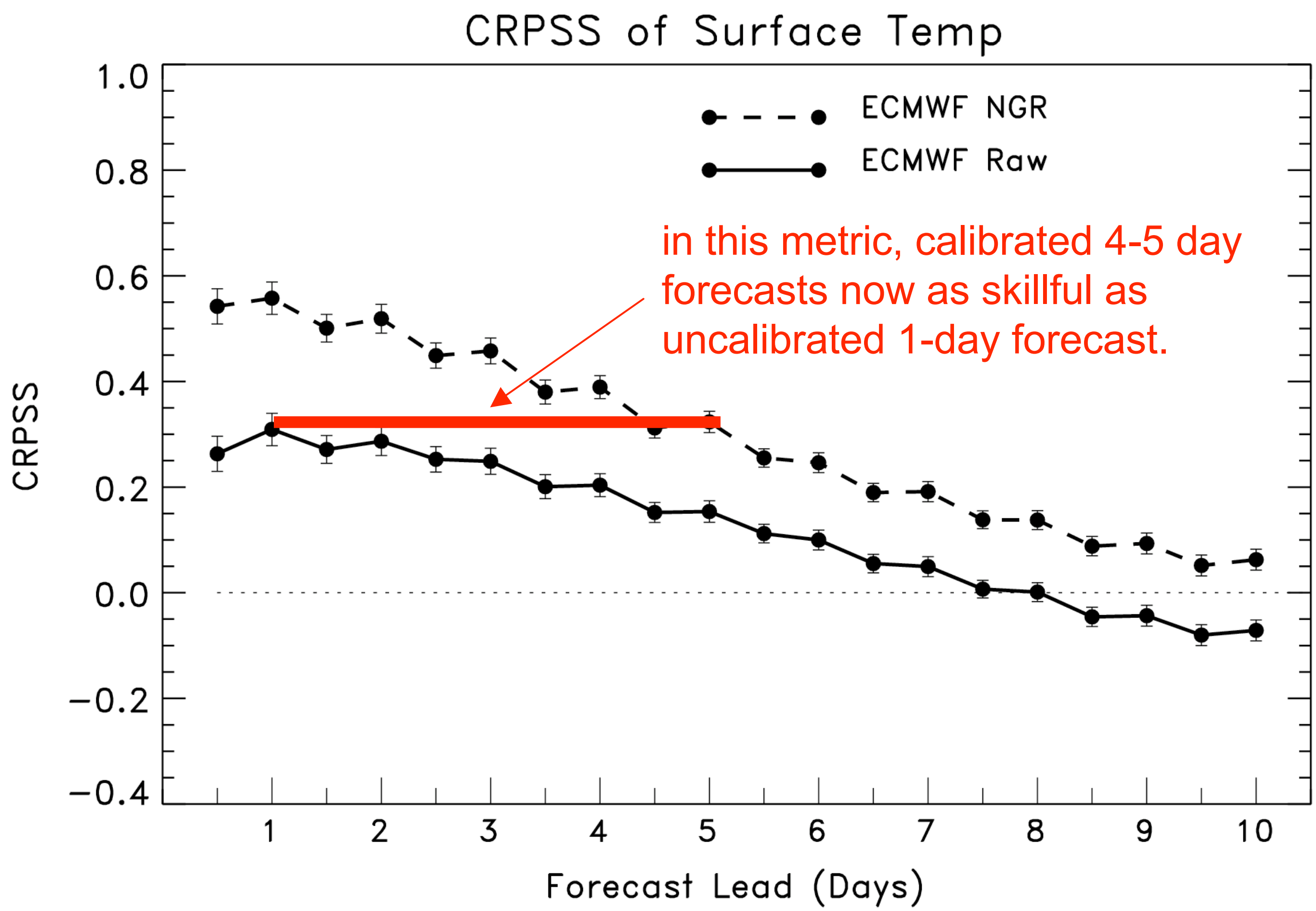
ECMWF, raw and post-processed



Small confidence intervals imply **significant improvement at all leads**

95% confidence interval using paired block bootstrap described in Hamill, *WAF*, 1999, **14**, p. 155

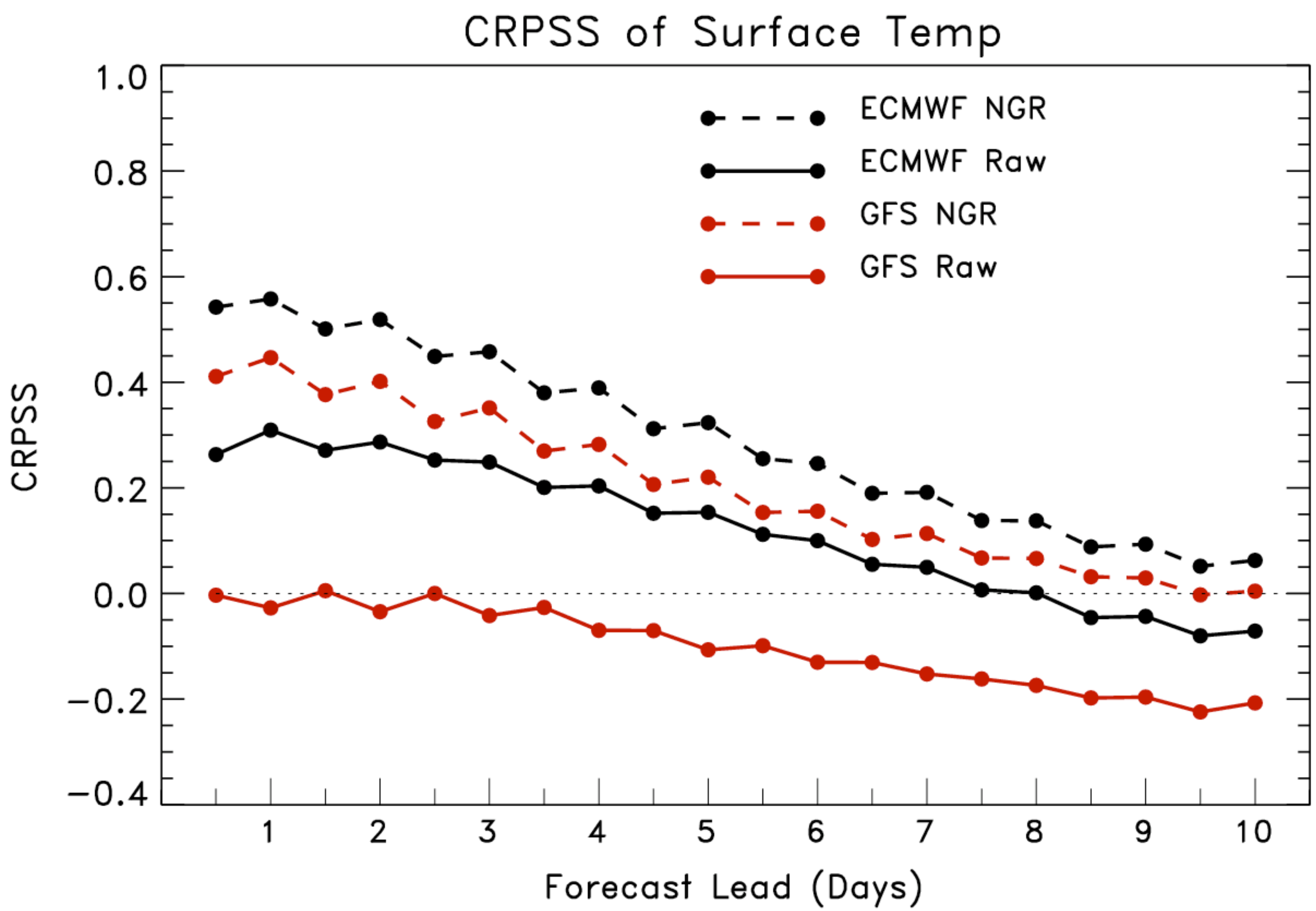
ECMWF, raw and post-processed



Small confidence intervals imply **significant improvement at all leads**

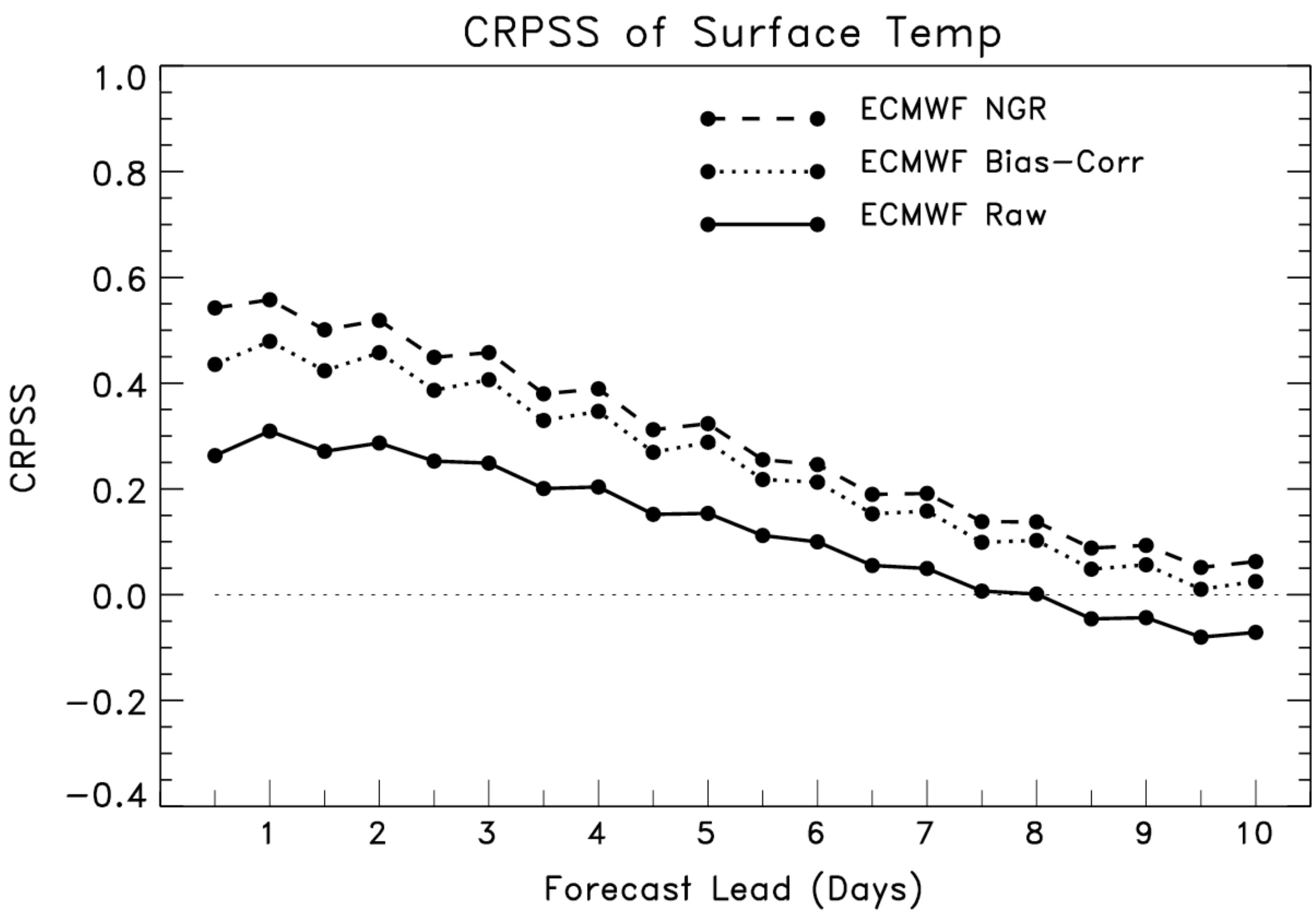
95% confidence interval using paired block bootstrap described in Hamill, *WAF*, 1999, **14**, p. 155

ECMWF and GFS



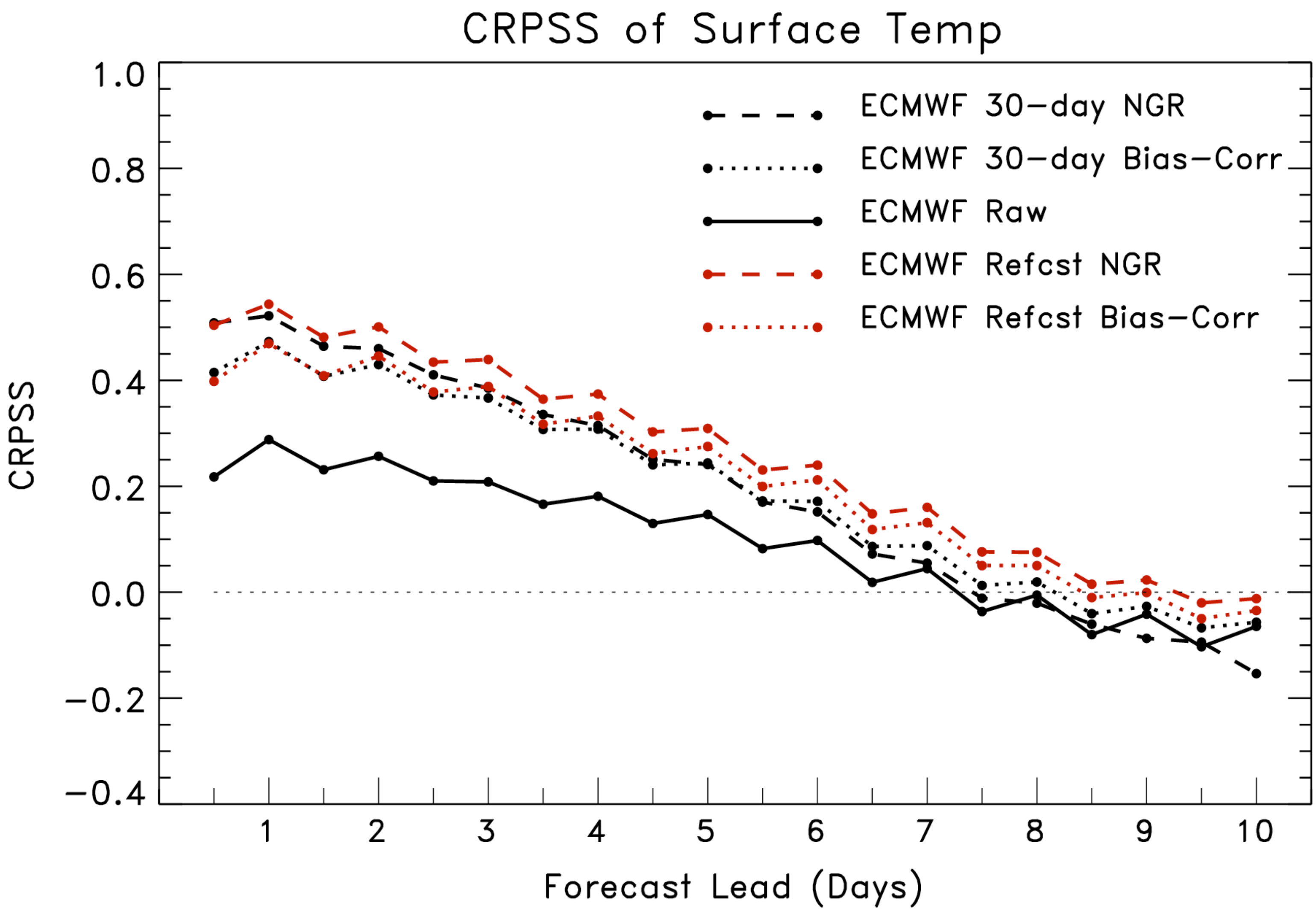
GFS reforecasts sub-sampled only to the dates of ECMWF reforecasts

How much from simple bias correction?

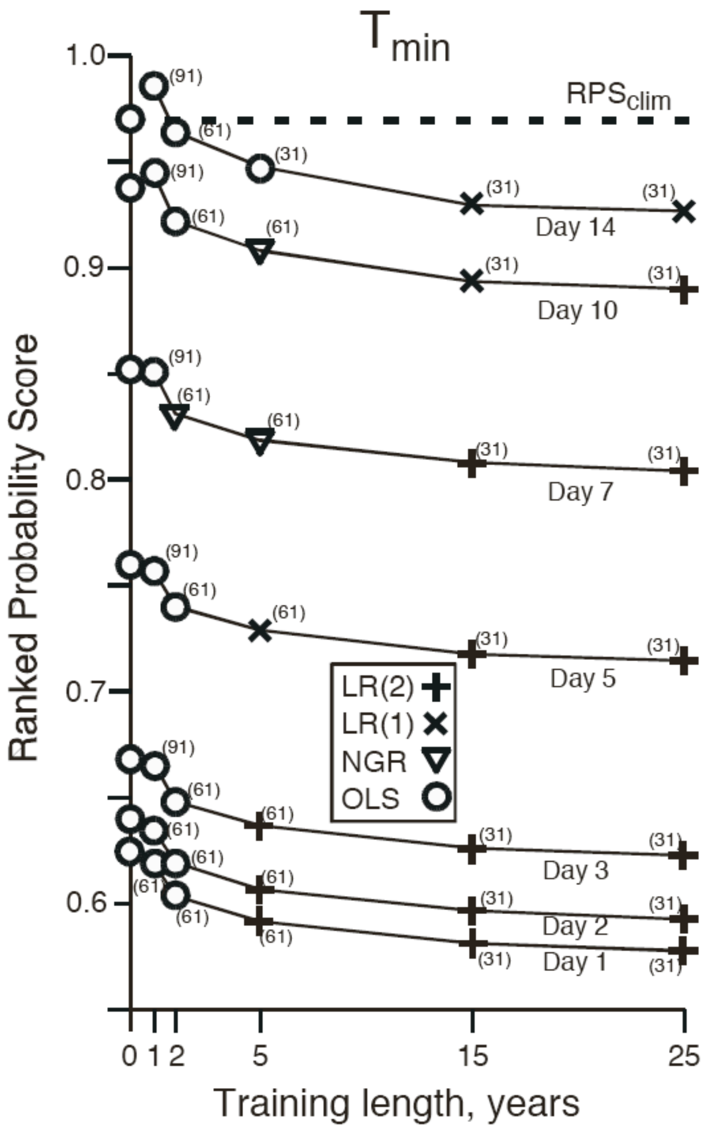


~ 60 percent of total improvement at short leads, 70 percent at longer leads.

How much from short training data sets?



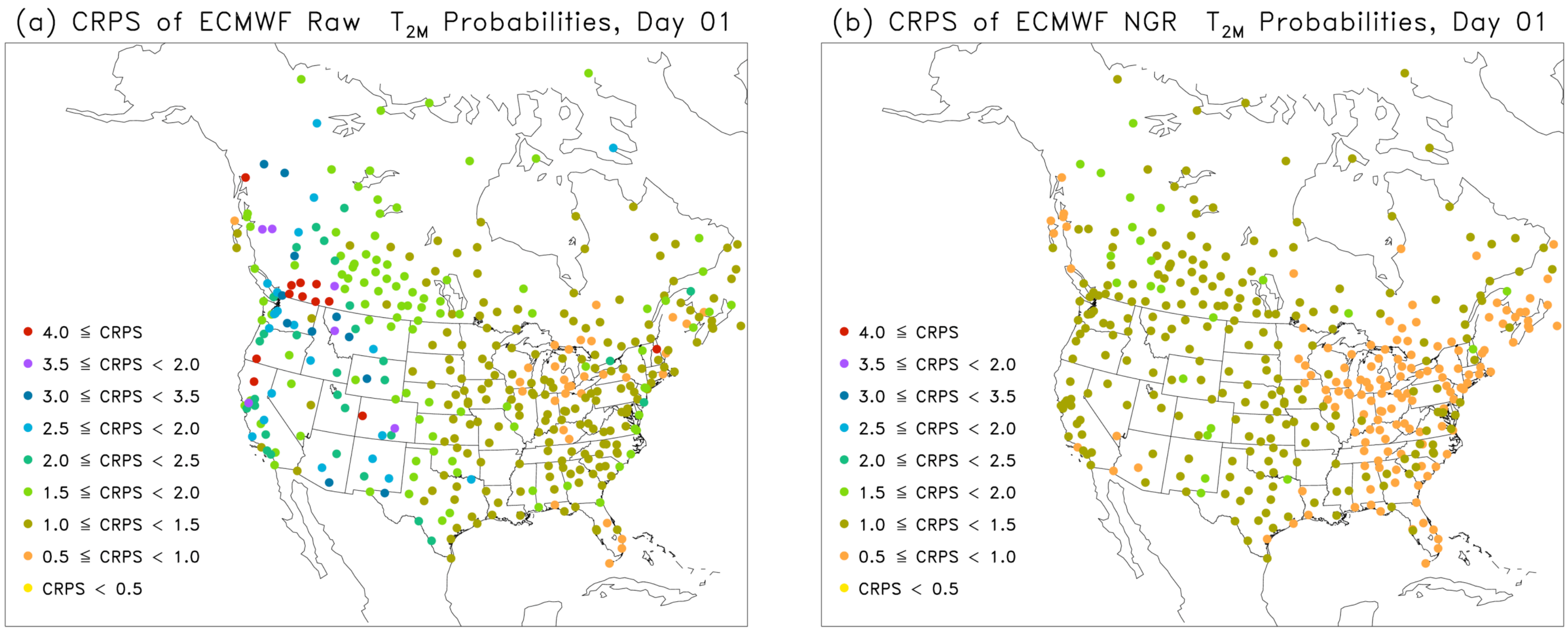
Results from GFS, T_{min}

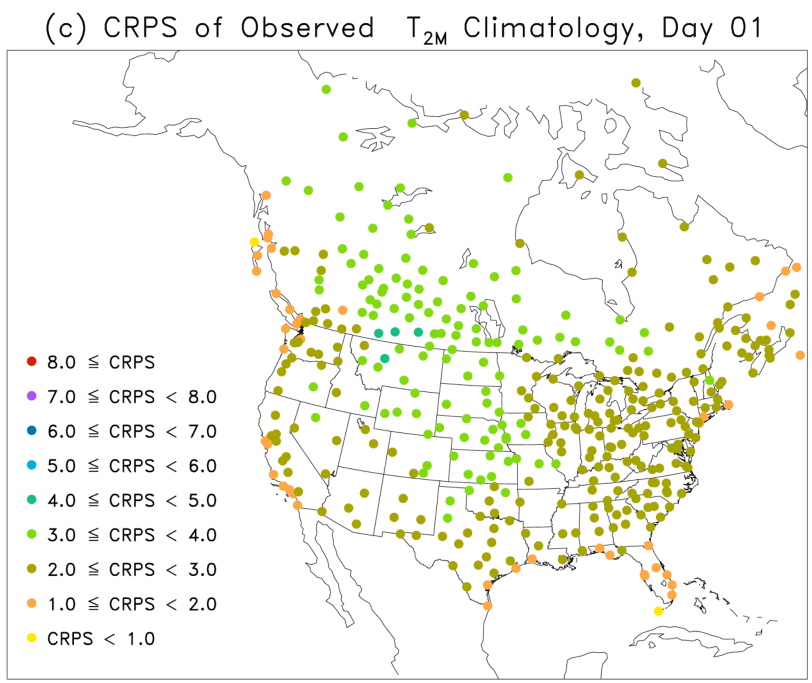


Are less optimistic results from short training data set with GFS reforecasts due to:

- poor model?
- use of full reforecast data set here, not subsample of 1x weekly?

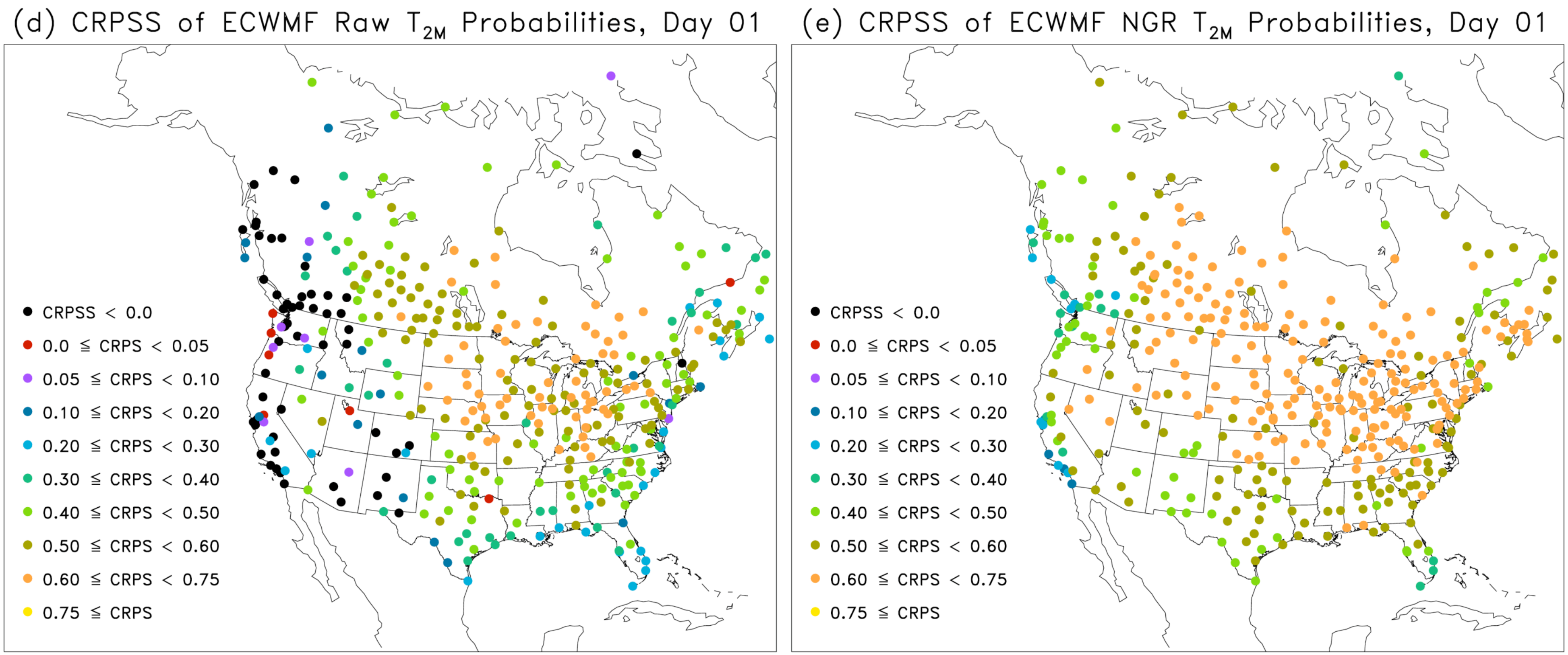
from Wilks and Hamill, 2006, Met. Apps., 13, 243-256.





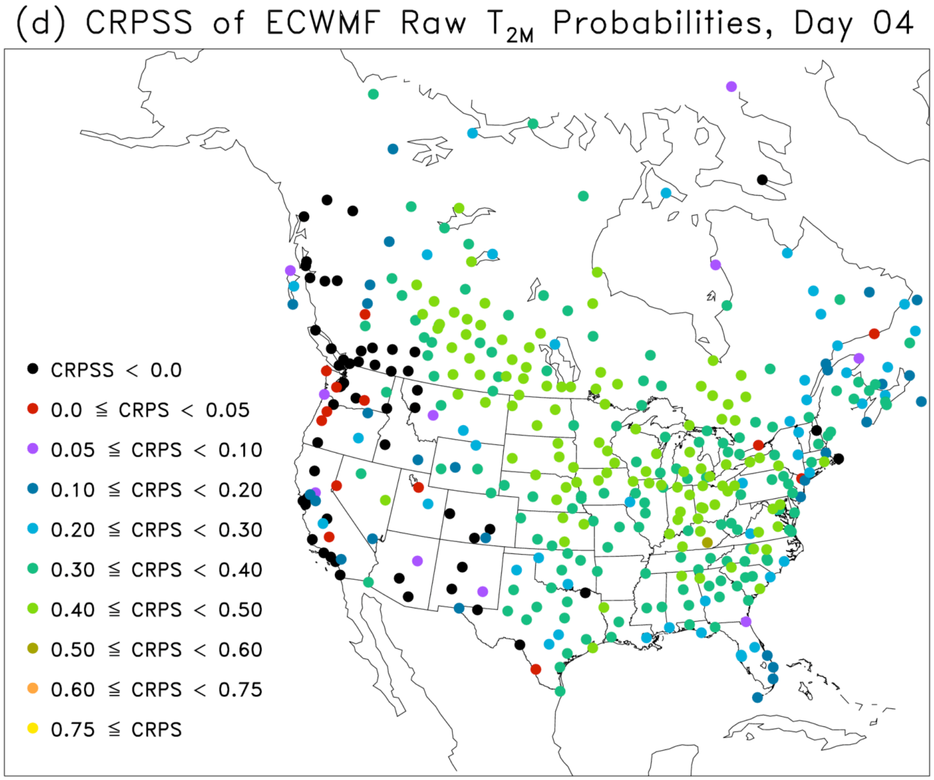
largest improvement at the stations with the highest original CRPS.

CRPSS, Day 1

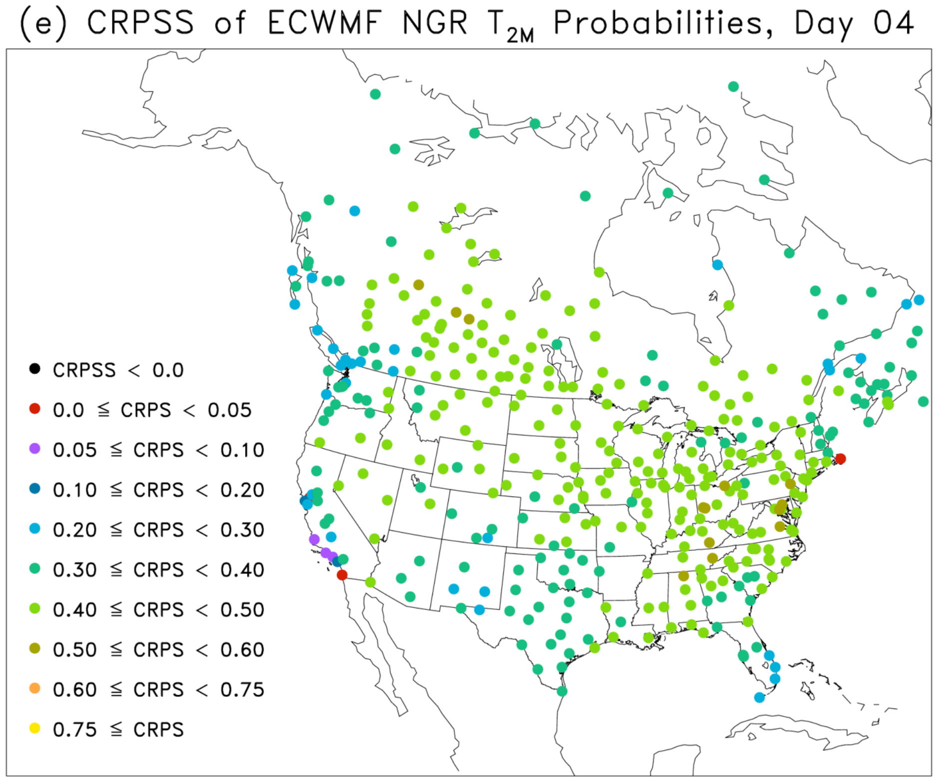


(d) CRPSS of ECWMF Raw T2M Probabilities, Day 01
CRPSS < 0.0
0.0 ≦ CRPS < 0.05
0.05 ≦ CRPS < 0.10
0.10 ≦ CRPS < 0.20
0.20 ≦ CRPS < 0.30
0.30 ≦ CRPS < 0.40
0.40 ≦ CRPS < 0.50
0.50 ≦ CRPS < 0.60
0.60 ≦ CRPS < 0.75
0.75 ≦ CRPS
(e) CRPSS of ECWMF NGR T2M Probabilities, Day 01
CRPSS < 0.0
0.0 ≦ CRPS < 0.05
0.05 ≦ CRPS < 0.10
0.10 ≦ CRPS < 0.20
0.20 ≦ CRPS < 0.30
0.30 ≦ CRPS < 0.40
0.40 ≦ CRPS < 0.50
0.50 ≦ CRPS < 0.60
0.60 ≦ CRPS < 0.75
0.75 ≦ CRPS

CRPSS, Day 4



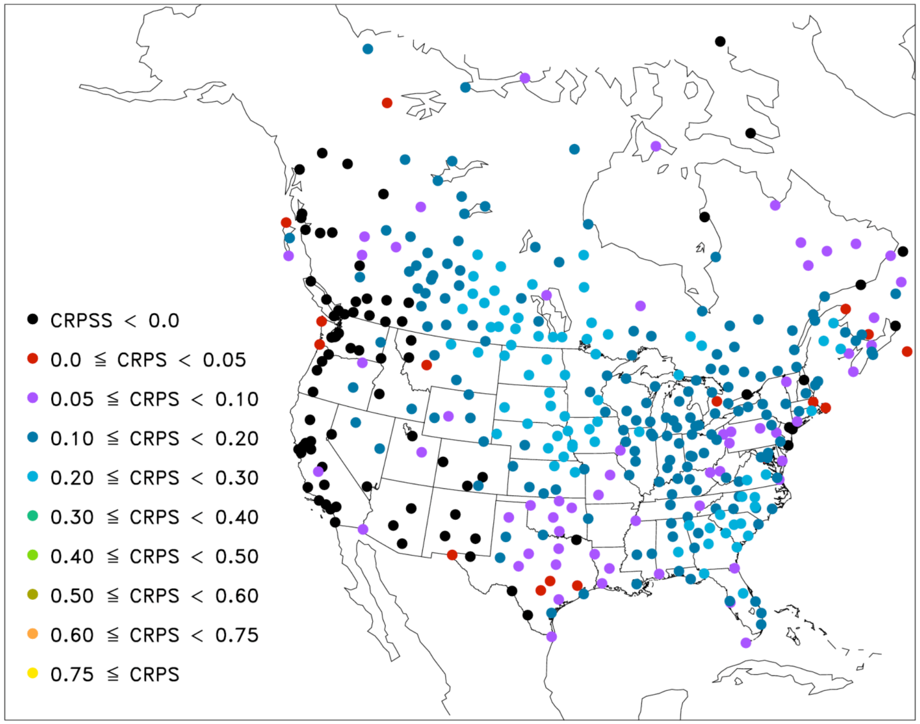
(d) CRPSS of ECWMF Raw T_{2M} Probabilities, Day 04
CRPSS < 0.0
0.0 ≤ CRPS < 0.05
0.05 ≤ CRPS < 0.10
0.10 ≤ CRPS < 0.20
0.20 ≤ CRPS < 0.30
0.30 ≤ CRPS < 0.40
0.40 ≤ CRPS < 0.50
0.50 ≤ CRPS < 0.60
0.60 ≤ CRPS < 0.75
0.75 ≤ CRPS



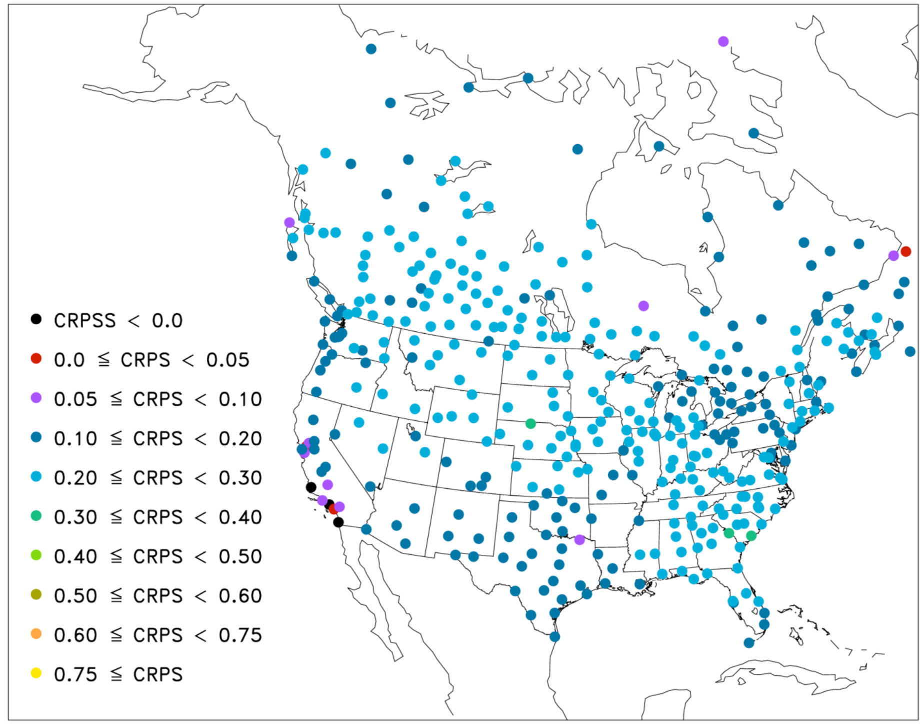
(e) CRPSS of ECWMF NGR T_{2M} Probabilities, Day 04
CRPSS < 0.0
0.0 ≤ CRPS < 0.05
0.05 ≤ CRPS < 0.10
0.10 ≤ CRPS < 0.20
0.20 ≤ CRPS < 0.30
0.30 ≤ CRPS < 0.40
0.40 ≤ CRPS < 0.50
0.50 ≤ CRPS < 0.60
0.60 ≤ CRPS < 0.75
0.75 ≤ CRPS

CRPSS, Day 7

(d) CRPSS of ECWMF Raw T_{2M} Probabilities, Day 07

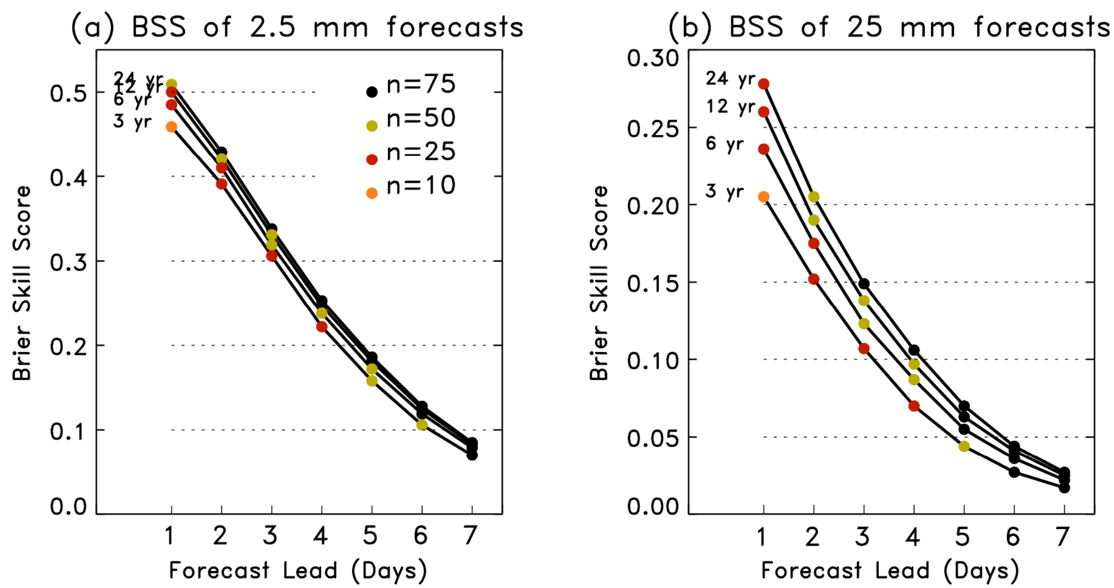


(e) CRPSS of ECWMF NGR T_{2M} Probabilities, Day 07



Notes

- Same benefit to precipitation calibration, winds, other variables? Perhaps not, w/o more full reforecast data set.



more rare events, like heavy precipitation forecasting, tend to benefit more from long reforecast data sets.

(results from full GFS reforecast data set. See Hamill et al., *BAMS*, January 2007)

Preliminary Conclusions

- Still substantial benefit to calibrating forecasts, even with a much better model than used in 1st-generation GFS reforecast.
- Old GFS + reforecast calibration » more skill than ECMWF uncalibrated.
- 30-day training does good job of calibration for short-term forecasts (consistent with previous NCEP results).
- Still need to test calibration of other variables (precipitation, wind speed, etc..)